DOCUMENT RESUME

ED 269 457 TM 860 310

AUTHOR Skaggs, Gary; Stevenson, Jose
TITLE A Comparison of ASCAL and LOGIST Parameter Estimation Programs.
PUB DATE Apr 86
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Algorithms; *Comparative Analysis; *Computer Software; *Estimation (Mathematics); Item Analysis; Item Banks; *Latent Trait Theory; *Maximum Likelihood Statistics; Programing; Test Construction
IDENTIFIERS *ASCAL Computer Program; Item Parameters; *LOGIST Computer Program; Three Parameter Model

ABSTRACT

This study assesses the accuracy of ASCAL, a microcomputer-based program for estimating item parameters for the three-parameter logistic model in item response theory. Item responses are generated from a three-parameter model, and item parameter estimates from ASCAL are compared to the generating item parameters and to estimates produced by LOGIST. ASCAL, part of an item banking/test construction software package known as MicroCAT, is described as a pseudo-Bayesian estimation process. The program, like LOGIST, is based on a pseudo-maximum likelihood algorithm. Unlike LOGIST, convergence is achieved by postulating prior distributions for examinee ability, item discrimination, and item lower asymptotes. Data are generated for sample sizes of 2,000 and 500, test lengths of 35 and 15, narrow and wide ranges of discrimination, and for centered and decentered ability distributions. The results show that the outputs from the two programs are usually more closely related to each other than either is to the original item parameters. Overall, both programs produce reasonable results for the larger data sets, but both have problems for small sets of data. It is concluded that if both programs are available, LOGIST would still probably be the program of choice due to its greater variety of usage options. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED269457

A COMPARISON OF ASCAL AND LOGIST PARAMETER ESTIMATION PROGRAMS

Gary Skaggs

Fairfax County (Va.) Public Schools

Jose Stevenson

Montgomery County (Md.) Public Schools

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. Skaggs

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, April 19, 1986.

TM 860 310

A COMPARISON OF ASCAL AND LOGIST PARAMETER ESTIMATION PROGRAMS

Abstract

The purpose of this study was to assess the accuracy of ASCAL, a microcomputer-based program for estimating item parameters for the three-parameter logistic model in item response theory (IRT). Item responses were generated from a three-parameter model, and item parameter estimates from ASCAL were compared to the generating item parameters and to estimates produced by LOGIST. Data were generated for samples sizes of 2,000 and 500, test lengths of 35 and 15, narrow and wide ranges of discrimination, and for centered and decentered ability distributions. The results showed that the outputs from the two programs were usually more closely related to each other than either was to the original item parameters. Overall, both programs produced reasonable results for the larger data sets, but both had problems for small sets of data.

A COMPARISON OF ASCAL AND LOGIST PARAMETER ESTIMATION PROGRAMS

Gary Skaggs

Fairfax Co. (Va.) Public Schools

Jose Stevenson

Montgomery Co. (Md.) Public Schools

The purpose of this paper is to compare two programs for estimating item parameters for a three-parameter logistic model based on item response theory (IRT). One of the programs, LOGIST V (Wingersky, Barton, & Lord, 1983), is currently the most widely used program for the three-parameter model and has been cited repeatedly in the literature. The other program, ASCAL, is a relative newcomer and part of an item banking/test construction software package known as MicroCAT (1984). The primary objective of this paper is to assess the accuracy of this new program by comparing its results with those of LOGIST for simulated under a variety of conditions.

Until recently, LOGIST was virtually the only program available for estimating parameters for the three-parameter model. However, several problems have been noted in its use. These include problems estimating discrimination and lower asymptote parameters (Lord, 1984; Baker, 1983), the need for large sample sizes and long tests (Reckase, 1979; Hulin, Lissak, & Drasgow, 1982), and considerable computer costs. LOGIST uses a maximum likelihood approach to estimate simultaneously item and

ability parameters. In recent years, alternate approaches have been developed, including marginal maximum likelihood (Mislevy & Bock, 1982) and Bayesian estimation (Swaminathan & Gifford, 1983). Still, these programs have not yet been widely investigated or accepted, and LOGIST continues to be the standard procedure for most IRT applications.

Another problem with using LOGIST, and most other IRT software, is that it is available only for certain mainframe computers. This eliminates many potential users of IRT methodology, in particular local school systems. Many software developers are recognizing this problem and are designing IRT programs for the microcomputer.

One software package that has been introduced to the market is MicroCAT (1984). Developed by Assessment Systems Corporation of Minneapolis, this package offers the potential to implement a wide range of IRT-based techniques. These include test construction, item banking, tailored testing, and parameter estimation. The estimation program, known as ASCAL, was written primarily by C. David Vale. The program, like LOGIST, is based on a pseudo-maximum likelihood algorithm. Unlike LOGIST, convergence is achieved by postulating prior distributions for examinee ability, item discrimination, and item lower asymptotes.

Since this program operates on an IBM-PC, it would have the potential of providing a valuable tool for anyone working in a setting where a mainframe computer is unavailable. This paper therefore will examine how well ASCAL recovers original item

parameters and how the program's estimates compare to those from LOGIST. These comparisons were done under a variety of conditions which would generally be considered less than optimal for accurate estimation.

OVERVIEW OF ASCAL PROGRAM

ASCAL is described as a "pseudo-Bayesian estimation process" (ASCAL, 1985, p.3). The program seeks to estimate item and ability parameters by maximizing a likelihood function into which Bayesian prior distributions have been added for ability, item discrimination, and item lower asymptote parameter estimates. The prior distribution for ability is a standard normal distribution. For the a and c parameters, the specified priors were symmetric beta distributions. The criterion likelihood function is shown below.

$$L^* = \prod_{i}^{N} \prod_{g}^{k_i} P_g(\theta)^{v_{gi}} Q_g(\theta)^{1-v_{gi}} \Phi(\theta_i) * f(a_g,r,s,j,k) * f(c_g,r,s,j,k)$$

where N is the number of examinees;

k_i is the number of items answered by examinee i;

$P_g = c_g + (1-c_g)[1+\exp\{-1.7a_g(\theta-b_g)\}]^{-1}$;

$Q_g = 1 - P_g$;

v_{gi} = 0 if item response is incorrect,
1 if item response is correct,
1/K otherwise, where K is the number of alternatives;

Φ is the standard normal density function;

$$f(x,r,s,j,k) = 1/B(r,s) * 1/(k-j) * \frac{x - j}{k - j}^{r-1} * 1 - \frac{x - j}{k - j}^{s-1}$$

and $B(r,s) = \int_0^1 y^{r-1} (1 - y)^{s-1} \, dy$

is the beta function, with upper and lower limits k and j, respectively, and beta function parameters r and s.

For the a parameter, r = s = 3.0 with upper and lower bounds of 2.6 and .3. For the c parameter, r = s = 5.0 with a lower bound of -.05 and an upper bound of .05 + 2/K.

To obtain parameter estimates, ASCAL employs a procedure similar to LOGIST where ability and item parameters are estimated separately. Initial ability estimates are based on raw formula scores. Initial item parameter estimates are computed from heuristic procedures described by Jensema (1976). Once abilities are estimated, they are grouped into 20 intervals in order to speed computations. Item parameters are then estimated until convergence is achieved. The sequence of ability estimation, ability grouping, and item parameter estimation is repeated until the item parameters converge or the maximum number of iterations has been reached.

One should not that ASCAL does not provide many of options and outputs available from LOGIST. For example, although omitted and not reached responses are permitted with ASCAL, no editing capabilities are included in the program. There is a Pearson chi-square model fit test for each item, but there are no

formulations for standard errors of estimates. One can specify the maximum number of iterations, but not convergence criteria or other aspects of the iteration process. Finally, because it is intended to be an item parameter estimation program, ability estimates are not output by ASCAL. Ability estimates can be obtained by an separate program within the MicroCAT software system.

METHOD

Generation of item responses

(0,1) item response data were generated from a three-parameter logistic model. In this model, the probability of a correct response to item i is a function of the item's difficulty (b_i), discrimination (a_i), and lower asymptote (c_i), and the examinee's ability (θ):

$$P_i(\theta) = c_i + (1-c_i)\{1 + \exp[-1.7a_i(\theta-b_i)]\}^{-1} \quad (1)$$

For a single examinee's response to one item, the probability in Equation was computed based on known item and ability parameters. This probability was then compared to a random number drawn from the (0,1) uniform distribution. If the probability was less than the random number, the item was scored as incorrect. Otherwise, the item was scored as correct. The GGUBS (IMSL, 1980) generator was used to generate random numbers.

Item and ability parameters

For all of the cases studied, item difficulties were uniformly distributed across a range of -2.0 to 2.0 logits. Lower asymptote parameters were uniformly distributed across a range of .1 to .3. There were four independent variables in this study: sample size, test length, range of item discrimination, and level of examinee ability.

Sample sizes of N=2,000 and 500 and test lengths of k=35 and 15 were simulated. The sample sizes were selected to be above and below the minimum recommended sample size for reasonable estimation (Reckase, 1979). The test lengths were chosen to be above and below the recommended minimum test length given the sample sizes that were studied (Hulin et al., 1983).

Item discriminations for all cases were centered at .8. In this study, two ranges of discrimination were examined. In the first case, discriminations ranged uniformly between .7 and .9. In the second case, the parameters ranged uniformly between .4 and 1.2. The first case resembles a situation in which items have been selected to have as nearly the same discrimination as possible. The second case represents a more typical calibration where there is a wider range of item discrimination. This type of manipulation of discrimination has not been usually done in Monte Carlo studies, but our preliminary work with LOGIST suggests that this may be an important factor in the accuracy of the estimates.

Finally, two levels of examinee ability were studied. In the first case, examinee ability was normally distributed with a mean of zero and standard deviation of one logit. In the second case, abilities were normally distributed with a mean of -1.0 and a standard deviation of one logit. Since the mean item difficulty is always zero, the first case is a situation where test difficulty is matched the ability of the examinees. In the second case, low ability examinees took a moderately difficult test. All abilities were generated using the GGNML (IMSL, 1980) generator.

Evaluation of estimation

A variety of descriptive statistics were computed to assess the accuracy of the item parameter estimates. For each of the generated data sets, LOGIST and ASCAL estimates were compared both to the original parameters and to each other. For a, b, and c parameters, correlations between estimates and original parameters were computed. In addition, the item parameter estimates from each program were transformed to the original scale by equalizing the means and standard deviations of item difficulties. Because ASCAL does not provide standard errors for the estimates, the transformations were not weighted in any way. For each parameter, mean differences (absolute values) were computed as well as the maximum deviation.

It is possible for item parameter values to differ across a specified range of ability while the item characteristic curves

differ only slightly (see Hulin et al., 1983). Because of this, the Root Mean Square Error (RMSE) index used in the Hulin study was also employed for this paper. This statistic has the following form:

$$RMSE = SQRT\ (1/201 \sum_{i=1}^{201} [\hat{P}_i(\theta_j) - P_i(\theta_j)]^2)$$

An RMSE was computed for each item. For a given case, both the mean and maximum RMSE were reported.

RESULTS

In running LOGIST, all of the default options were used, except that ability grouping was employed in order to speed up computations. Convergence was typically achieved in twenty to twenty five stages using up to a minute and a half of computer processing time (about $50). For ASCAL, the larger data sets took about one hour to run and the smaller ones about 30 minutes, using the math coprocessor chip.

Item Difficulty

Table 1 shows descriptive statistics for item difficulty estimates from the two programs. All of the correlations were .90 or better. Most were .96 or higher, indicating that even for smaller data sets, the item difficulty estimates from both programs were rank ordered in a similar manner to the original item difficulties and to each other. The mean absolute differences were consistently small for larger samples and longer

tests. When sample size and/or test length were reduced, maximum differences varied considerably across the cases. For combinations of a short test and a small sample, the mean differences for ASCAL were smaller than those for LOGIST. Occasionally, for either program large maximum differences appeared. In these cases, the mean difference was likewise larger to some degree, indicating that outliers greatly influenced the magnitude of these indices.

Item Discrimination

In Table 2, the same descriptive statistics are shown for item discrimination. Correlations ranged from -.01 to .91. Not surprisingly, the lowest correlations occurred for both programs in cases where the range of initial item discrimination was narrow. For these cases, one or two outliers substantially reduced the magnitude of the correlations. For wider ranges of discrimination, the correlations are typical of those observed for LOGIST in the literature (Ree, 1979; Hulin et al., 1983; Qualls & Ansley, 1985). In fourteen of the sixteen cases, LOGIST and ASCAL estimates correlated more highly with each other than they did with the original estimates, indicating that similar types of bias for item discrimination were found with both programs.

An examination of mean differences suggests that the accuracy of the estimates declined more as a function of test length than of sample size. This seemed to be more of a factor

for ASCAL which had average mean differences of about .5 for the cases of short test lengths as opposed to roughly .4 for LOGIST. The maximum differences of both programs varied considerably across all cases, but were comparable to each other.

Both programs produced discrimination estimates that were positively skewed even original discriminations were uniformly distributed. Estimates also tended to have a wider range than the original parameters, having a lower bound similar to that of the original range, but with a much high upper limit. In many cases, one or two of the estimates were at the upper limit allowed by the program, 2.0 for LOGIST and 2.5 for ASCAL.

Item Lower Asymptotes

Table 3 reports descriptive statistics for estimating lower asymptote parameters. As might be expected, correlations varied considerably, ranging from -.46 to .86. As in the case of the correlations reported earlier, the values were extremely susceptible to outliers. Since the initial range for this parameter was quite narrow (.2), the correlations were surprisingly high. As in the case of item discrimination, the estimates from the two programs tended to correlate more highly with each other than with the original parameters.

The values for mean and maximum absolute differences indicate a high degree of invariance to the conditions imposed by this study. Mean differences ranged from .03 to .09 for both LOGIST and ASCAL. These results indicate that both programs

estimate the c parameter with the same degree of accuracy and that these estimates do not deteriorate for data sets as small as the ones used in this study.

Recovery of ICC's

In Table 4, RMSE's are presented comparing the two programs to the original ICC's and to each other. The mean values were also plotted in Figures 1 and 2 in order to aid in interpretation. Figures 1a and 1b plots mean RMSE's for cases where a narrow range of discrimination was used. Figures 2a and 2b show mean RMSE's for wide ranges of discrimination.

ENTER FIGURES 1 AND 2 ABOUT HERE

For all sixteen cases, the mean RMSE's for LOGIST were smaller for ability distributions centered at zero logits than for the decentered distributions ($\bar{\theta} = -1.0$). For ASCAL, this result held only for longer tests. For the shorter tests, smaller RMSE's were found for decentered distributions in three out four cases. In two of the cases involving shorter tests, the mean RMSE was unexpectedly larger for larger samples. This result was likely due to sampling fluctuations than to inconsistency of the estimates. Because the means were based on only three values, one large RMSE (i.e. poor estimates for one item) could greatly inflate the mean. Practically speaking, differences in mean RMSE's of .01 for the shorter tests are

probably insignificant. The mean RMSE's for the longer tests appeared to be more consistent.

Mean RMSE's were lower for ASCAL than for LOGIST in ten of the sixteen cases. For larger sample sizes and longer tests, the programs were comparable for centered ability distributions, but ASCAL mean RMSE's were larger for decentered distributions. For larger samples and shorter tests, the programs produced similar results for decentered distributions, but LOGIST was superior for centered distributions. For smaller samples, ASCAL RMSE's were lower than those for LOGIST in six of eight cases. The only exceptions were for centered distributions with a narrow range of discrimination. Finally, when RMSE's were computed comparing ASCAL and LOGIST ICC's, the mean RMSE was lower than those for either program in thirteen of the sixteen cases. This supports previously reported results that the two programs tend to produce estimates more similar to each other than to the original item parameters.

DISCUSSION

Of all the descriptive statistics used in this study, the RMSE index seemed to provide the clearest picture of the accuracy of the estimates from the two programs, supporting results reported by Hulin et al. (1983). However, all the indices used here involved summations across items. For shorter tests in particular, all indices suffered from the effects of outliers,

one or two poorly estimated items. Correlations seemed to be the most susceptible.

In general, both programs provided estimates of item parameters that were very similar to each other. The accuracy of both programs suffered when sample size and/or test length was reduced. In general, ASCAL produced more accurate estimates for smaller sample sizes, but seemed to be less accurate with decentered ability distributions with the larger data sets.

Both programs had some difficulty estimating item discriminations. These problems seemed to be intensified more for shorter tests than for smaller samples. Despite the fact that original discriminations were uniformly distributed, both ASCAL and LOGIST produced estimates that were positively skewed and distributed across a wider range than the original parameters.

The range of item discrimination and the level of examinee ability relative to test difficulty are not variables typically used in studies examining estimation programs. The results of this study show that estimates from LOGIST and ASCAL are sensitive to these variables and that these variables interact in complex ways with test length and sample size.

The variety of cases studied here is certainly not exhaustive of the wide range of types of data sets that may be found in practice, but these cases do represent a reasonable range of conditions under which LOGIST was known vary in its accuracy. For the situations simulated here, ASCAL seems to

provide slightly better estimates than LOGIST for small data sets. For conditions under which LOGIST is known to perform well, ASCAL provides comparable estimates.

Based on the results of this study, ASCAL provides a reasonable alternative to LOGIST. It produces estimates similar to LOGIST. Its estimates are not as accurate in situations where LOGIST has likewise been found to have difficulty. If both programs are available, LOGIST would still probably be the program of choice due to its greater variety of usage options. If LOGIST is unavailable, ASCAL can be used with some confidence in its accuracy for larger data sets such as those studied in this paper.

REFERENCES

Assessment Systems Corporation. (1984) User's manual for the MicroCAT testing system. St. Paul, MN: Author.

Baker, F.B. (1983) Comparison of ability metrics obtained under two latent trait theory procedures. Applied Psychological Measurement, 7, 97-110.

Hulin, C.L., Lissak, R.I., & Drasgow, F. (1982) Recovery of two- and three-parameter logistic item characteristic curves: a Monte Carlo study. Applied Psychological Measurement, 6, 249-260.

IMSL, Inc. (1980) International mathematical and statistical libraries reference manual. Houston, TX: IMSL, Inc.

Jensema, C.J. (1976) A simple technique for estimating latent trait mental test parameters. Educational and psychological measurement, 36, 705-715.

Lord, F.M. (1984) Maximum likelihood and bayesian parameter estimation in item response theory. (ETS Technical Report No. RR-84-30-ONR). Princeton, NJ: Educational Testing Service.

Mislevy, R.J. & Bock, R.D. (1982) BILOG - Maximum likelihood item analysis and test scoring: LOGISTIC model. Chicago: International Educational Services.

Qualls, A.L. & Ansley, T.N. (1985) A comparison of item and ability parameter estimates derived from LOGIST and BILOG. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Ree, M.J. (1979) Estimating item characteristic curves. Applied Psychological Measurement, 3, 371-385.

Reckase, M.D. (1979) Item pool construction for use with latent trait models. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Swaminathan, H. & Gifford, J.A. (1983) Estimation of parameters in the three-parameter. In D. Weiss (ed.), New Horizons in Testing. New York: Academic Press.

Wingersky, M.S., Barton, M.H., & Lord, F.M. (1983) LOGIST: a computer program for estimating examinee ability and item characteristic curve parameters. LOGIST 5, Version 1. Princeton, NJ: Educational Testing Service.

Table 1

Correlations, Mean Absolute Differences, and Maximum Absolute Differences for Item Difficulty

N	k	range of a	$\bar{\theta}$	r_{ol}[1]	r_{oa}[2]	r_{al}[3]	M_{oa}[4]	M_{ol}	Max_{oa}[5]	Max_{ol}
2,000	35	.2	0	.99	.99	.99	.12	.13	.36	.47
2,000	35	.2	-1	.99	.99	.99	.13	.14	.44	.54
2,000	35	.8	0	.99	.99	.99	.11	.14	.41	.41
2,000	35	.8	-1	.99	.99	.99	.13	.13	.49	.46
2,000	15	.2	0	.99	.99	.99	.14	.11	.28	.27
2,000	15	.2	-1	.98	.99	.98	.12	.22	.25	.50
2,000	15	.8	0	.99	.98	.99	.20	.15	.43	.32
2,000	15	.8	-1	.99	.99	.99	.15	.17	.47	.49
500	35	.2	0	.98	.96	.97	.23	.20	1.47	.55
500	35	.2	-1	.97	.98	.99	.18	.24	.57	.55
500	35	.8	0	.96	.98	.98	.19	.28	.65	1.09
500	35	.8	-1	.93	.95	.96	.27	.32	1.19	1.41
500	15	.2	0	.97	.98	.99	.18	.24	.68	.52
500	15	.2	-1	.96	.99	.98	.15	.27	.55	.75
500	15	.8	0	.96	.99	.97	.13	.23	.43	.97
500	15	.8	-1	.90	.96	.94	.25	.40	.87	1.36

---

[1] original versus LOGIST parameters

[2] original versus ASCAL parameters

[3] ASCAL versus LOGIST parameter estimates

[4] M = Mean absolute difference

[5] Max = Maximum absolute difference

Table 2

Correlations, Mean Absolute Differences, and Maximum Absolute Differences for Item Discrimination

N	k	range of a	$\bar{\theta}$	r_{ol}[1]	r_{oa}[2]	r_{al}[3]	M_{oa}[4]	M_{ol}	Max_{oa}[5]	Max_{ol}
2,000	35	.2	0	.29	.40	.90	.14	.13	.50	.69
2,000	35	.2	-1	.29	.44	.75	.22	.21	1.04	1.27
2,000	35	.8	0	.89	.81	.93	.18	.14	1.10	.64
2,000	35	.8	-1	.79	.64	.79	.19	.15	.86	.71
2,000	15	.2	0	.48	.16	.60	.41	.29	1.54	.76
2,000	15	.2	-1	.01	.44	.45	.62	.47	1.75	1.48
2,000	15	.8	0	.91	.89	.86	.57	.22	1.42	.50
2,000	15	.8	-1	.77	.72	.80	.52	.31	1.59	.87
500	35	.2	0	-.01	.05	.77	.24	.31	.90	1.20
500	35	.2	-1	.15	.15	.89	.27	.33	1.28	1.27
500	35	.8	0	.56	.65	.87	.25	.30	1.05	1.02
500	35	.8	-1	.17	.44	.61	.32	.33	1.52	1.59
500	15	.2	0	.63	.30	.84	.63	.49	1.65	1.34
500	15	.2	-1	.35	.77	.55	.48	.41	1.28	1.25
500	15	.8	0	.64	.76	.88	.41	.50	1.11	1.33
500	15	.8	-1	.62	.48	.87	.55	.70	1.63	1.63

---

[1] original versus LOGIST parameters

[2] original versus ASCAL parameters

[3] ASCAL versus LOGIST parameter estimates

[4] M = Mean absolute difference

[5] Max = Maximum absolute difference

Table 3

Correlations, Mean Absolute Differences, and Maximum Absolute Differences for Item Lower Asymptotes

N	k	range of a	$\bar{\theta}$	r_{ol}[1]	r_{oa}[2]	r_{al}[3]	M_{oa}[4]	M_{ol}	Max_{oa}[5]	Max_{ol}
2,000	35	.2	0	.37	.50	.75	.04	.05	.13	.14
2,000	35	.2	-1	.58	.47	.86	.05	.04	.22	.15
2,000	35	.8	0	.49	.64	.80	.04	.05	.20	.16
2,000	35	.8	-1	.54	.54	.77	.05	.05	.20	.15
2,000	15	.2	0	.33	.24	.79	.08	.06	.16	.15
2,000	15	.2	-1	.59	.39	.84	.07	.04	.20	.21
2,000	15	.8	0	.01	-.46	.53	.09	.07	.30	.15
2,000	15	.8	-1	.46	.69	.82	.05	.06	.14	.20
500	35	.2	0	.52	.45	.92	.05	.06	.16	.20
500	35	.2	-1	.40	.52	.83	.05	.08	.16	.19
500	35	.8	0	.16	.42	.68	.05	.09	.15	.28
500	35	.8	-1	.27	.33	.61	.05	.09	.16	.30
500	15	.2	0	.38	.09	.30	.07	.07	.20	.18
500	15	.2	-1	.51	.52	.64	.05	.08	.19	.21
500	15	.8	0	.37	.39	.60	.06	.09	.17	.24
500	15	.8	-1	.01	.28	.48	.07	.08	.21	.24

---

[1] original versus LOGIST parameters

[2] original versus ASCAL parameters

[3] ASCAL versus LOGIST parameter estimates

[4] M = Mean absolute difference

[5] Max = Maximum absolute difference

Table 4

Root Mean Square Errors

				Original/ASCAL			Original/LOGIST			ASCAL/LOGIST		
N	k	range or a	$\bar{\theta}$	Mean	S.D.	Max	Mean	S.D.	Max	Mean	S.D.	Max
2,000	35	.2	0	.026	.012	.062	.029	.014	.069	.013	.008	.033
2,000	35	.2	-1	.041	.025	.114	.035	.024	.119	.023	.023	.134
2,000	35	.8	0	.028	.016	.086	.032	.015	.078	.016	.009	.037
2,000	35	.8	-1	.040	.022	.097	.033	.012	.063	.021	.019	.110
2,000	15	.2	0	.056	.022	.100	.048	.015	.067	.027	.017	.061
2,000	15	.2	-1	.055	.039	.169	.064	.028	.115	.058	.024	.115
2,000	15	.8	0	.060	.041	.165	.043	.018	.073	.050	.024	.108
2,000	15	.8	-1	.050	.034	.121	.055	.018	.096	.037	.019	.080
500	35	.2	0	.046	.037	.238	.044	.019	.078	.025	.041	.257
500	35	.2	-1	.043	.023	.099	.054	.025	.115	.028	.023	.113
500	35	.8	0	.038	.020	.102	.046	.021	.105	.027	.016	.062
500	35	.8	-1	.053	.046	.248	.063	.041	.178	.045	.044	.214
500	15	.2	0	.066	.043	.166	.053	.033	.122	.042	.034	.158
500	15	.2	-1	.051	.027	.110	.070	.032	.144	.061	.031	.149
500	15	.8	0	.055	.021	.102	.085	.039	.148	.055	.023	.090
500	15	.8	-1	.060	.032	.128	.099	.044	.200	.073	.041	.162

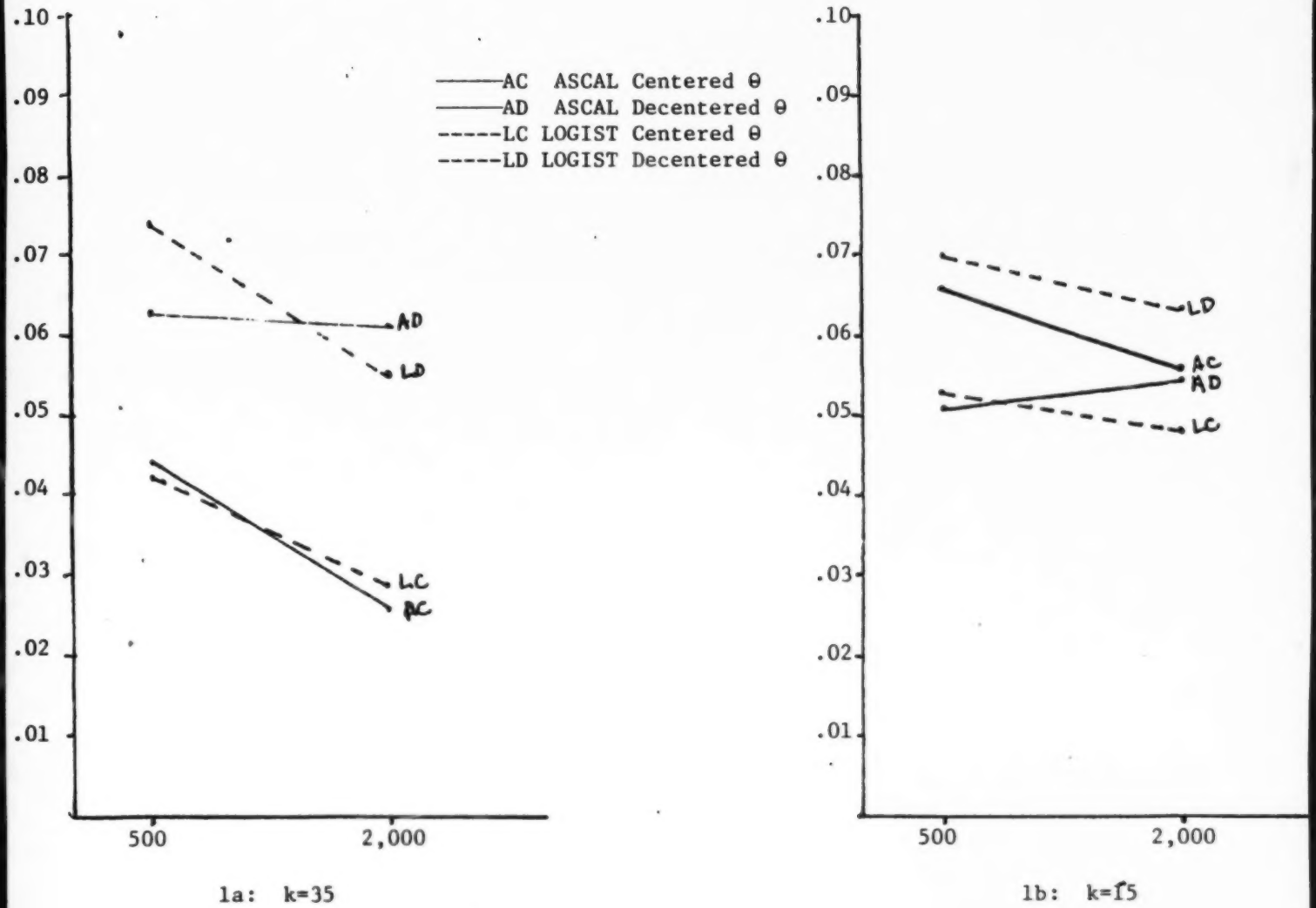

Figure 1. Plots of mean RMSE's where a's range from .7 to .9.

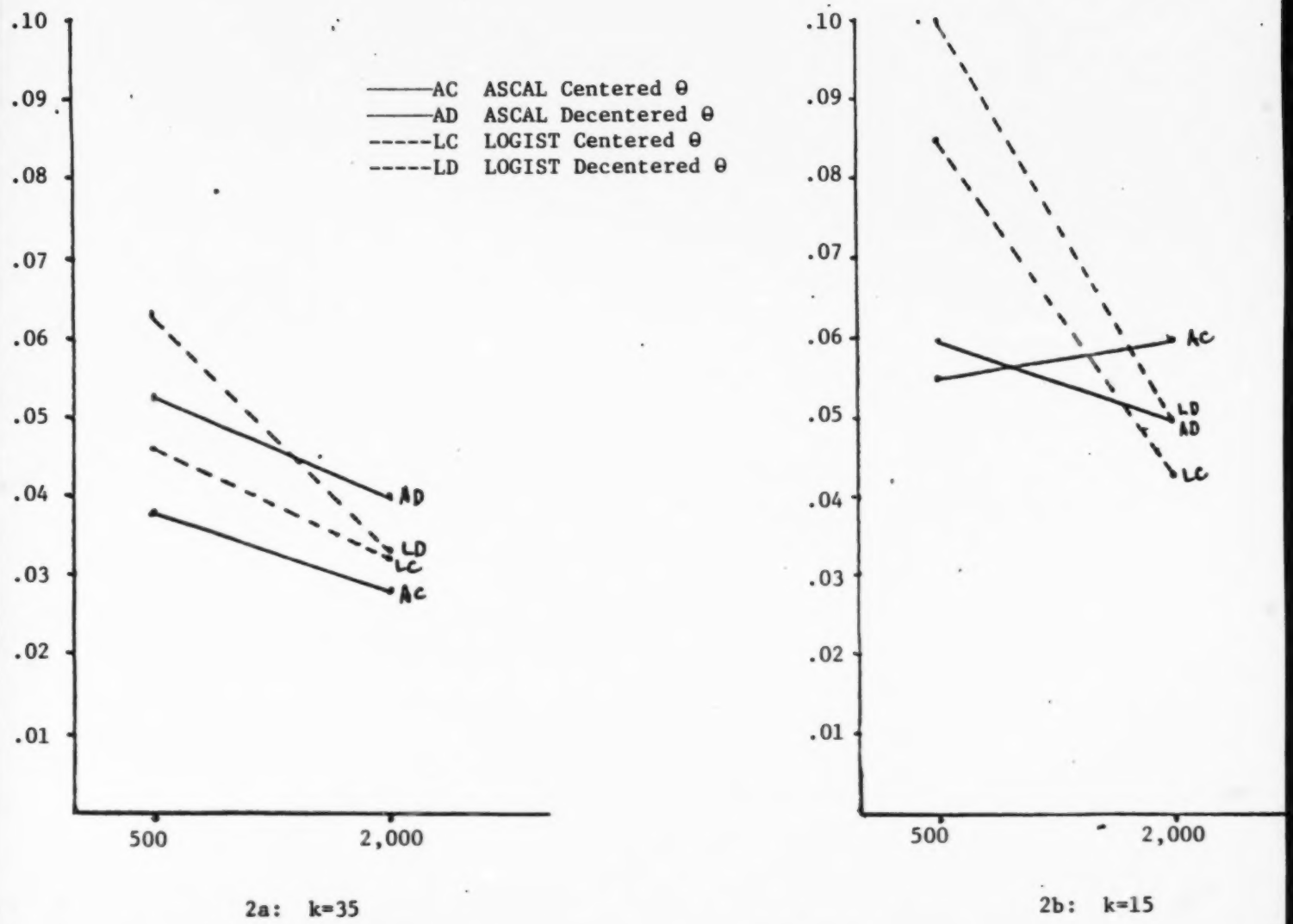

Figure 2. Plots of mean RMSE's where a's range from .4 to 1.2.